In the early 1970s the University of California-Berkeley was sued for gender discrimination in its graduate admissions. The lawsuit alleged that female applicants were unfairly admitted at a lower rate compared to male applicants.

University departments generally make their own admissions decisions. The following fictionalized contingency table represents admissions decisions broken down by gender for six departments at Berkeley in the early 1970s.

	Whether Admitted			
	Male		Female	
Department	Yes	No	Yes	No
Business Administration	512	313	89	19
Physics	353	207	17	8
History	120	205	202	391
English	138	279	131	244
Psychology	53	138	94	299
Philosophy	22	351	24	317

(**Note:** This data set is based on real admissions data from Berkeley in the early 1970s. Some actual numbers are identical, but I have made up the department labels for the purposes of this exam–*Derek*.)

Problem 1

Fit a model relating admissions decisions (A) to gender (G) and department (D). Select a model that is **not saturated**. Specifically, do the following:

- Provide the code used to fit the model. (**10 points**)
- Display the summary of the model fit. (**2 points**)
- Write down the equation of the model you estimated, defining all quantities appropriately. (**5 points**)

Problem 2

Investigate how well the model fits the data. Specifically address the following:

- Use the model you fit in Problem 1 to calculate the expected counts for each cell in the contingency table. Display and label these counts. (**6 points**)
- Are there particular cells where the model appears to fit badly? If so, discuss what may be going wrong. (**Note:** if you choose, you may formally calculate model residuals using the methods outlined in subsection 5.2.1 of Bilder and Loughin. But this is not necessary! An informal comparison of observed and expected counts suffices.) (**6 points**)
- Carry out a formal goodness-of-fit test to assess the model fit. Interpret the results in context. (**4 points**)

Problem 3

Using a good-fitting model, investigate the effect of gender G on admissions decisions A. Specifically, do the following:

- If your first model did not pass the goodness-of-fit test in Problem 2, then estimate a model which provides a better fit to the data.
  - Provide the code used to fit this new model. (**8 points**)
  - Display a summary of the model. (**2 points**)
  - Write down the equation of the model you estimated, defining all terms. (**5 points**)
  - Establish that the model fits the data better. (**4 points**)
- Using the good-fitting model you ultimately choose, estimate all AG conditional odds ratios (conditional on department D).
  - Provide the code used for the calculations. (**6 points**)
  - Report all the estimates and interpret them in context. (**4 points**)
- Calculate confidence intervals for all AG conditional odds ratios. (It is **not** necessary to adjust for multiple confidence intervals, but you may do so if you wish.)
  - Specify the confidence level. (**2 points**)
  - Provide the code used for the calculations. (**8 points**)
  - Report all results and interpret them in context. (**6 points**)
- Does gender have an impact on departments' admissions decisions? Discuss. (**6 points**)

Problem 4

Recall that Berkeley was sued for gender bias against women in its graduate admissions. In light of this, address the following questions:

- Is Simpson's paradox occurring in this contingency table? If yes, provide some reasons for why it is occurring, using the data. (**10 points**)
- Why do you think Berkeley was sued for gender discrimination in its graduate admissions? Based on your investigation of the data, do you think the lawsuit has merit? (**6 points**)